Full title: Gene expression patterns predictive of radiation-enhanced colon tumorigenesis: Diet as a countermeasure

Final Report

PI: Joanne R. Lupton
Co-I: Nancy D. Turner (Co-PI)
Robert S. Chapkin
Leslie A. Braby
John R. Ford
Raymond J. Carroll

Period of Report: 5/16/2003 – 4/30/2005

Institution:
Texas A&M University
Faculty of Nutrition
2253 TAMU
College Station, TX 77843-2253

Grant number: NAG 9-1523

Date: 6/22/2005

Introduction

There is currently no reliable, non-invasive method to determine the response to radiation exposure and development of colon cancer throughout the disease process. One potential method to monitor changes in colonocyte gene expression that occur during the various stages of the disease is through global transcriptional profiling using exfoliated colonocyte mRNA.

This proof of principle project has as its main goal the modification of existing protocols in order to improve the effectiveness of polyA isolation from feces in preparation for analysis of gene expression using microarray techniques.

Specific Aims

#1. The first aim was to optimize polyA isolation and sample handling steps prior to array hybridization in order to develop a protocol that could reliably be used to measure gene expression from exfoliated colonocytes.
#2. The second aim was to establish normalization procedures capable of handling the problem of sample degradation that occurs in fecal samples.

The CPR question most applicable to this research is 31d "How can effective biomarkers of carcinogenic risk from space radiation be developed and validated?" The work conducted addresses this question through the development of protocols that would enable biomarker identification for radiation exposure and colon carcinogenesis. Routine screening throughout long-term missions will be critical if the monitoring of radiation exposure and carcinogenesis induction is to be successful. A non-invasive procedure would make sampling easier and less stressful on the astronauts. Therefore, a technique using sloughed cells to determine the exposure to radiation and the response to that exposure, as measured by changes in gene expression, would permit routine screening. This research did not identify any new CPR questions.

The risks to personnel in space from naturally occurring radiations are generally considered to be one of the most serious limitations to long duration human space missions (BEIR, 1990; BEIR, 1998; Williams et al., 1999). Of all the risks of radiation exposure, the highest priority is development of cancer. Although definitive epidemiological data are not available on which types of cancer may be enhanced by radiation exposure (due to low numbers of astronauts compared to numbers required for statistical significance), one study using pilots shows an added risk for colorectal cancer (Rodionov and Potapov, 2002). This is not surprising since colon cancer is the second leading cause of death from cancer in the US today and strikes men and women nearly equally (Landis et al., 1998). Colon cancer can be induced by chemical carcinogens or by radiation, but the combination of the two appears to be at least additive or possibly synergistic. Fortunately, of all the cancers, colon cancer is the most amenable to dietary intervention and one of the easier to cure if detected early in the disease process. We had already developed a noninvasive technique to analyze changes in colonocyte gene expression from colon cells exfoliated into fecal material through standard PCR techniques (Davidson et al., 1995, 1998, 2003). The goal of this work was to improve upon the existing protocol so that the isolated material could be used for global transcriptional profiling over time to document radiation exposure and colon cancer development in a non-invasive manner.

Materials & Methods

Weanling Sprague-Dawley rats (n = 60) were provided one of two diets that contained either the combination of fish oil and pectin, or the combination of corn oil and cellulose. Half of the rats were exposed to 1 Gy of 1 GeV/nucleon Fe ions at Brookhaven National Laboratory 3 weeks after starting the experimental diets. All rats received injections of a colon-specific carcinogen (azoxymethane, AOM) on days corresponding to 10 and 17 after irradiation.

Fecal samples were collected 7, 14 and 28 weeks after the last AOM injection. Fecal pellets were collected soon after defecation, stored in RNA protection buffer and frozen (-80°C) until processed. A sample handling flowchart demonstrates the steps involved in sample handling after collection of the fecal sample and the steps that were modified are identified by a "*".

Sample Handling Flowchart

Fecal Sample

↓ *Isolate/ *RT Reaction

cDNA generated

↓ *PCR

Amplified DNA

↓ *In vitro transcription

Biotin labeled cRNA

↓ Hybridization

Analyze array

Modifications to Isolation Procedures

Modifications to our previously published methods included development of a new polyA isolation method, improved method of sample quality assessment, and a new technique for generation of biotin labeled cRNA for hybridization on microarrays. These changes are the subject of a patent application we are developing and, as such, will not be explained in enabling detail in this report. The net effect of these changes are an increased yield of polyA, an increase in RNA quality used for analysis and an RNA template that is suitable for generating biotin labeled cRNA for microarray hybridization using GE CodeLink Rat Genome Arrays.

Microarray Data Normalization Procedures

A new normalization procedure was needed for these data because the results obtained using standard normalization procedures (Global median and Quantile) were not informative. The existing procedures were not capable of handling sample degradation that occurs while the cells are in transit in the large intestine. Between the degradation that occurs upon cells being

sloughed from the epithelial layer and fermentation by bacteria, samples collected from the fecal stream will always be partially degraded. One of the existing techniques, Global median normalization, was not able to eliminate nonlinear trends in comparisons between arrays within a diet (Figure 1). Another frequently used method, the Quantile method, removed the normal individual gene-to-gene correspondence between arrays within the same treatment (Figure 2). The outcomes generated by these existing procedures would be biased and uninformative. Therefore, we developed a new two-stage normalization procedure to deal with fecal microarray data. A description of the procedure and the justification for it follows and is the subject of a paper by Liu et al. (2005) submitted to the journal Bioinformatics (see appendix).

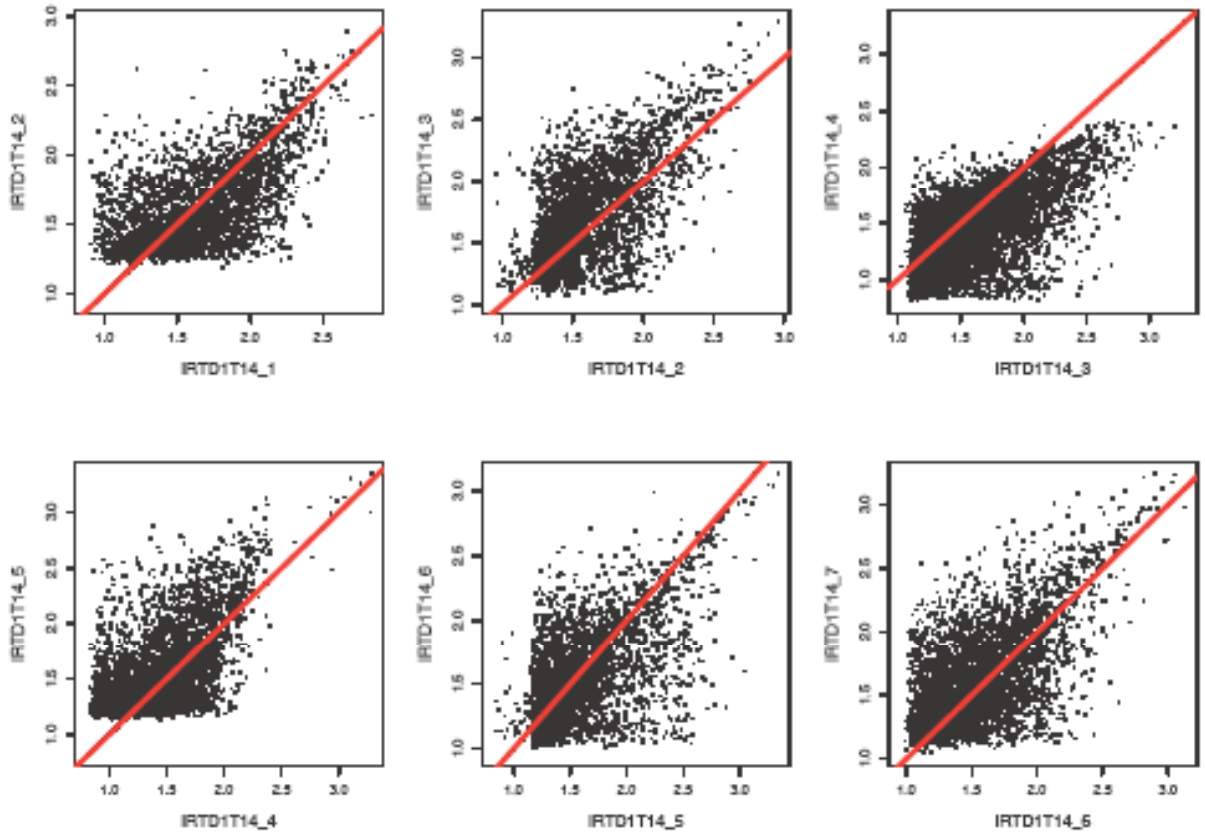

Figure 1. Scatter plots between two different Diet 1 arrays normalized by the Global median method. This method does not eliminate nonlinear trends in comparisons between arrays (demonstrated by uneven distribution along 45-degree line).

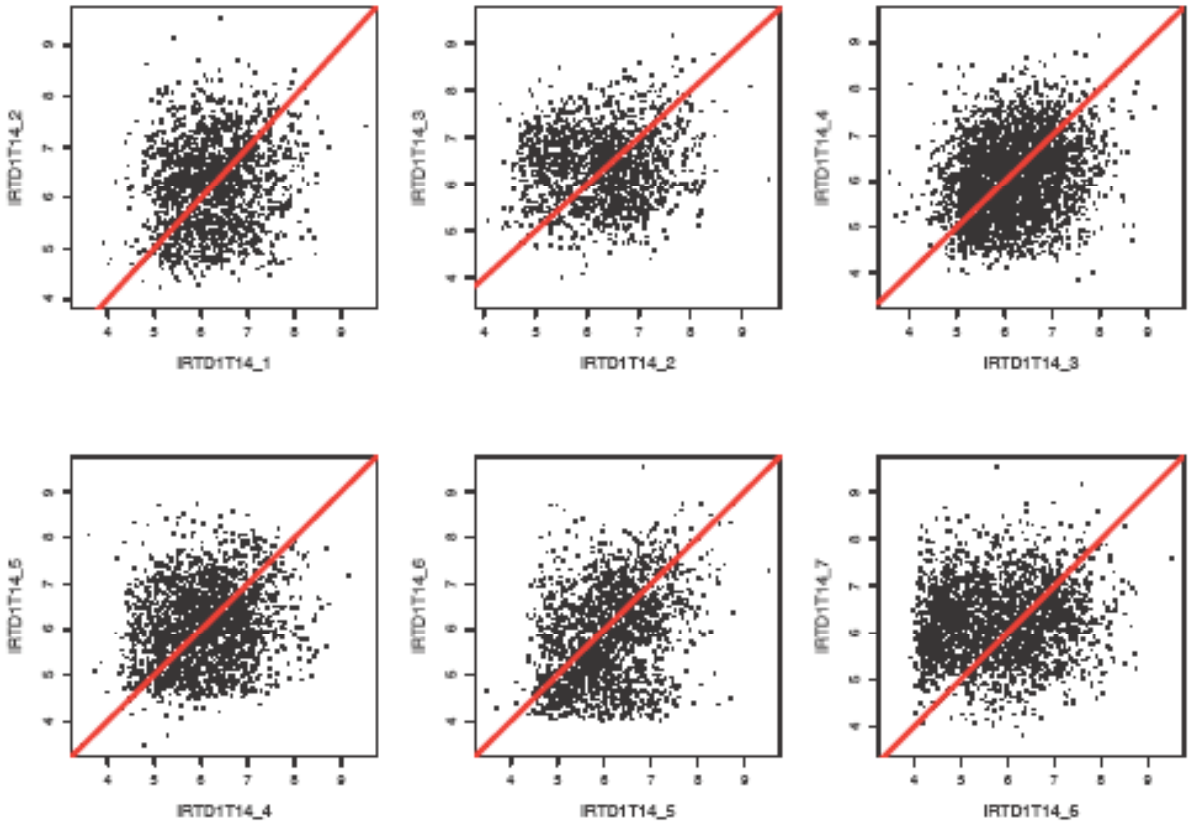

Figure 2. Scatter plots between two different Diet 1 arrays normalized using the Quantile method. This procedure removed the normal individual gene-to-gene correspondence, which results in the widely scattered pattern in these graphs instead of the normal distribution along the 45-degree line.

The basic principles behind the new two-stage normalization procedure include the following assumptions:

• We assume that even with different proportions of probes with low intensities (L) in each array, there exists a central "peak" within the probes with good intensities (G), such that the location and variation embedded in these central peaks enable us to perform a location-scale transformation. The transformation allows us to roughly align at least the central part of the majority of the arrays within the same treatment.

• Besides "aligning" all arrays through location-scale transformations, we used local smoothing techniques to correct for intensity-based biases within each treatment group. This correction procedure mimics the idea behind non-linear transformation and corrects upward/downward biases of each array by comparison to the trend of its peer(s), as discussed by Yang et al. (2002). For CodeLink arrays, the peers are the arrays within the same treatment group, since there are no green/red channels. A simple robust inclusion step is introduced so that the majority of the "L"

probes were excluded during this process of identifying nonlinear transformation to align signals in arrays within the same treatment group.

A complete description of the technique can be found in the submitted paper, which is included in the appendix. Briefly, we do the following.

The first step of the procedure is to linearly transform the "central" 25% of the log-2 transformed gene expression levels in each array, to best match the central 25% of the standard normal distribution. The central location is defined as the model of the central peak among the good probes. With the objective of regularizing the good probes, we performed a simple gene-by-gene analysis in order to include the "L" probes that behave similarly to the "G" probes for the same gene (a robust inclusion step).

The second step of the procedure involves using the local smoothing method to correct for any intensity-based biases associated with measurements in arrays with the same treatment group. We establish a baseline array for each treatment group, which is the median of the "qualified" measurements for each gene. The objective is to then correct any shared upward or downward non-linear trends by individual genes from the same array, as compared to the baseline array. For example, if one array whose up-regulated genes consistently have higher values compared to its peers in the same treatment group, it will be identified by an upward trend compared to the baseline median array. This intensity-based bias can then be removed. We have tried both local polynomial fits to the conditional mean and conditional median, where the latter estimation is carried out using local polynomial quantile regression ("quantreg" in R) described by Koenker (2005).

This transformation puts the emphasis on within treatment group alignment. When the mis-alignment among different treatment groups is a concern, a very simple modification using quantile transformation can be added. For the majority of cases, if a gene has all "L" spots in one treatment group, it tends to have all "L" spots in all treatment groups. This is because bacteria may target the breakdown of certain nucleotide sequences. This simple step warrants that the probe medians from different treatment groups follow the same distribution. This is a much weaker assumption than requiring that the usable probe values in each bioarray follow the same distribution.

Repeatability

Twenty of the samples, distributed throughout all treatment x time combinations, were selected to run duplicate arrays in order to assess repeatability. Typically repeatability in microarray analyses only means repeating hybridization from a single source of labeled cRNA. However, our samples were taken from the isolated polyA and run through all of the procedures. Therefore, repeatability measurements in our experiment document the variation inherent in all aspects of sample handling and processing, except for the initial isolation step.

Results

New polyA isolation results in higher yield

The new polyA isolation method, which is based on a commercial kit, has the advantage of yielding a significantly greater amount of polyA RNA than our earlier published methods. From a typical fecal sample, approximately 10-125 ng of polyA RNA can be obtained per gram feces. In contrast, our previous method (Davidson et al., 1998) resulted in a yield of only 1-10 ng polyA RNA/g feces, using a similar starting sample (Table 1).

Table 1. Yield of polyA from fecal isolations

Method	Range of polyA yield	Average polyA yield
Previous method	1-10 ng/g feces	4.4 ± 0.9 ng/g feces
Current method	10-125 ng/g feces	96 ± 15.5 ng/g feces

New polyA isolation method results in higher quality

The quality of the resultant polyA RNA is improved, based on Agilent Bioanalyzer analysis. The Agilent Bioanalyzer can assess RNA quality using the elution profile of nucleic acid species moving through a microcapillary chip. Previously, it was nearly impossible to determine quality of polyA RNA preparations. Therefore, the advent of this technology has allowed us to assess the quality of our samples prior to further processing. Since fecal RNA is inherently subject to degradation during passage through the gastrointestinal tract, sample integrity can be compromised. These poor quality samples can be eliminated from the experiment stream based on the results of this analysis. Examples of a typically acceptable polyA sample and an unacceptable sample are shown in Figure 3. Upon extensive degradation, the peak shifts to the left (elutes faster) and is composed of primarily small RNA fragments.

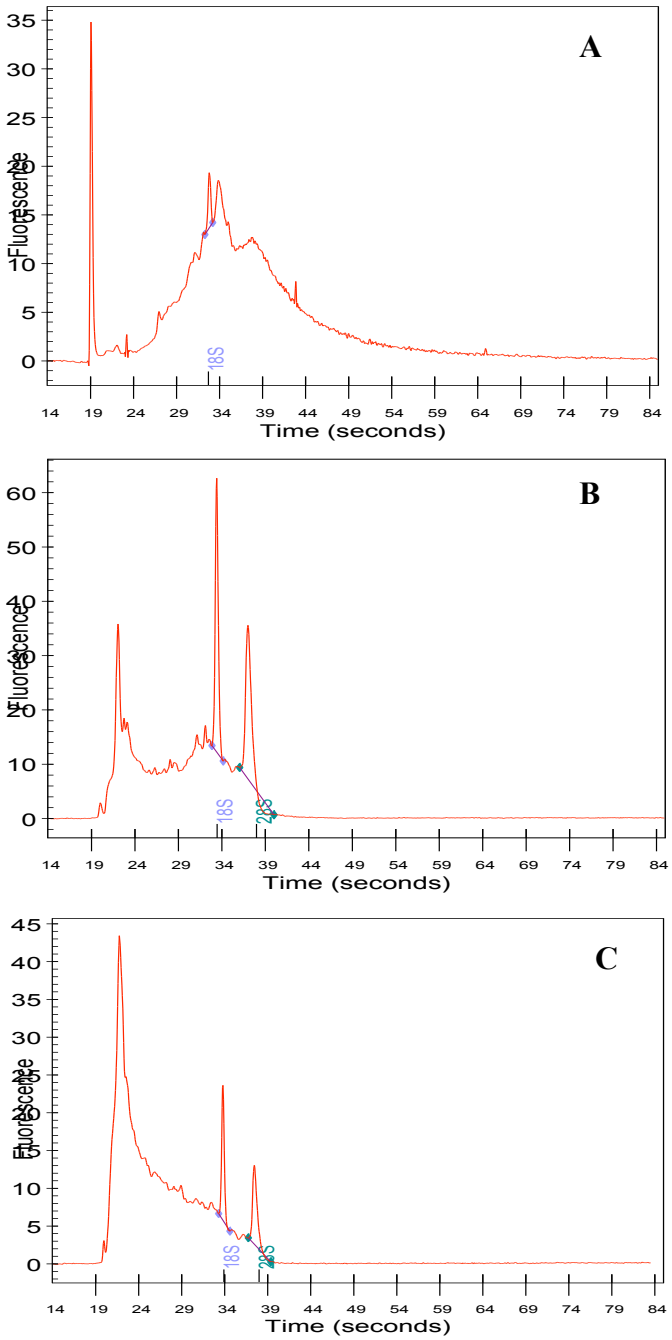

Figure 3. Agilent Bioanalyzer profiles of (A) commercial polyA RNA, (B) acceptable fecal polyA, and (C) unacceptable degraded fecal polyA.

Development of protocol for cRNA production

A method for generating biotin-labeled cRNA was developed in order to hybridize to CodeLink Rat Genome microarrays. Commercial kits would not produce adequate amounts of labeled cRNA to generate acceptable signals on the arrays. Therefore, we developed a specialized protocol for the fecal polyA samples based on an amplification step followed by an in vitro transcription reaction incorporating biotin labeled nucleotides. Starting with 5 ng fecal polyA, we were able to generate >10 μg labeled cRNA for hybridization to the CodeLink arrays. Hybridization and processing of the arrays followed the standard CodeLink protocol (Amersham).

Results of the New Data Normalization Procedure

We applied our method to the data generated from the CodeLink fecal microarrays. In step I, we matched the "central" 25% of the G probes to the equivalent central 25% standard normal distribution. Both the local polynomial mean and median estimates were carried out in the second step to adjust intensity-based biases among arrays within the same treatment.

In Figure 4, gene-by-gene scatter plots between 2 different Diet 1 arrays resulting from the two-stage normalization procedure are shown. The points are appropriately scattered around the 45-degree lines. Unlike the results of Global median transformation in Figure 1, the nonlinear trends have been properly removed. Differing from observations after Quantile transformations, Figure 4 shows that after the two-stage transformation, if a gene is highly expressed in one array, it tends to have a larger chance to be highly expressed in another array within the same treatment group.

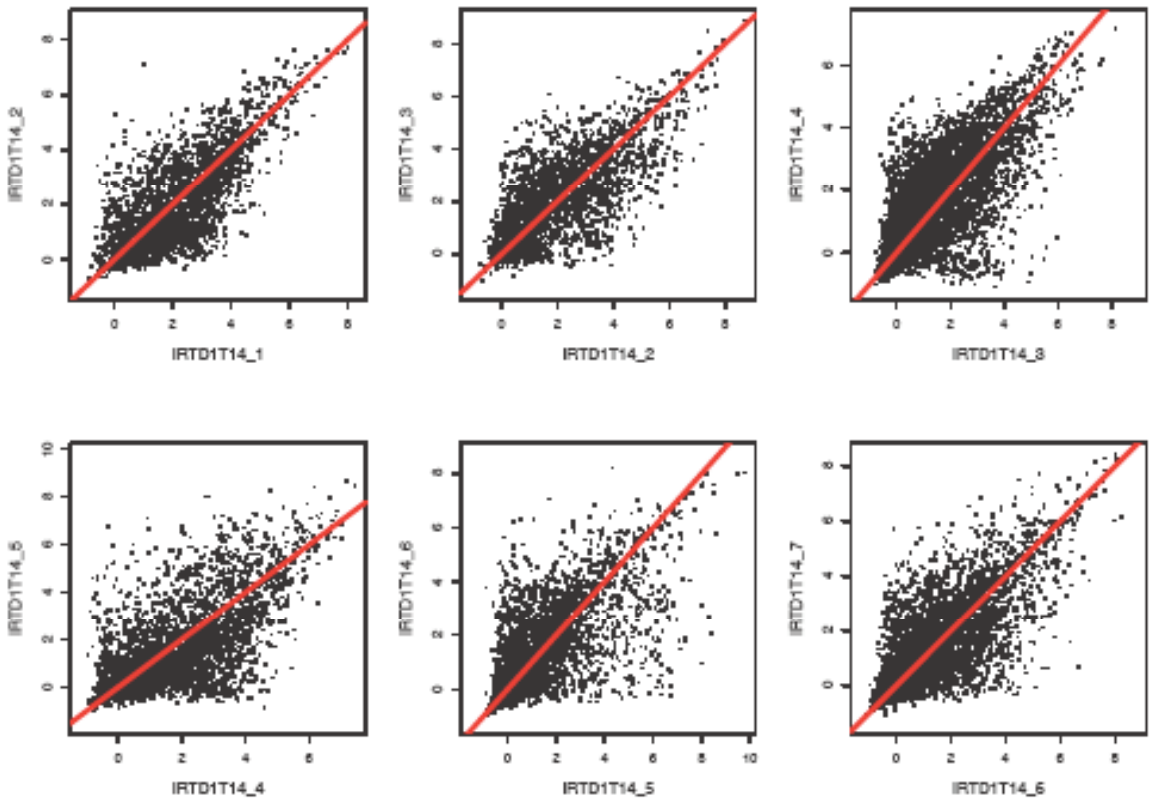

Figure 4. Scatter plots between two different Diet 1 arrays normalized using our new Two-stage procedure. The new procedure does a better job of producing the expected distribution along the 45-degree line.

Repeatability

Correlation coefficients for the duplicated arrays ranged from 0.63 to 0.96. The mean of all correlations was 0.77 with a standard error of 0.004. This degree of agreement between the two separate runs of the samples indicates that, although there are problems of sample degradation, the samples can reliably yield highly repeatable results.

Gene Expression

Distinct populations of genes were influenced by radiation exposure at 7, 14 and 28 weeks. The pattern of changes over these times will likely cluster within specific pathways, which is the subject of continuing work. Of interest, though is that even 28 weeks after the last carcinogen injection, which was 17 days after radiation exposure, that there were many different genes still being differentially expressed between the rats exposed to radiation and those that were not. For example, several cytochrome P450 genes (CYP4B1, CYP4A3) and cytochrome c oxidase subunit genes (COX5b, subunit IV) were altered by radiation exposure. These observations suggest that radiation, even at this late stage of tumorigenesis, was still altering basic cell functions dealing with substrate oxidation and the production and handling of reactive intermediates.

Summary and Conclusions

Modifications to the isolation and processing procedures for fecal polyA samples has resulted in highly improved quality and quantity of polyA that can be used for micorarray analysis of gene expression. We have developed a two-stage normalization procedure to process the data generated from the partially degraded RNA that is capable of handling this type of data, yet can be computed in a reasonable amount of time. Analysis of the data indicates there are strong effects of radiation on gene expression documented throughout this model of colon carcinogenesis, and that the effects can be tracked using a non-invasive procedure.

References

BEIR, V. Health Effect of Exposure to Low Levels of Ionizing Radiation: BEIR VII. N. R. C. Committee on Health Effects of Ionizing Radiation. Washington, DC (1998).

BEIR, V. Health Effects of Low Levels of Ionizing Radiation: BEIR V. N. R. C. Committee on the Biological Effects of Ionizing Radiation. Washington, DC (1990).

Davidson, L.A., Aymond, C.M., Jiang, Y-H., Turner, N.D., Lupton, J.R. and Chapkin, R.S. Non-invasive detection of fecal protein kinase C □II and □ messenger RNA: putative biomarkers for colon cancer. Carcinogenesis 19: 253-257 (1998).

Davidson, L.A., Jiang, Y.-H., Lupton, J.R. and Chapkin, R.S. Noninvasive detection of putative biomarkers for colon cancer using fecal messenger RNA. Cancer Epidemiology, Biomarkers & Prevention 4:643-647 (1995).

Davidson, L.A., Lupton, J.R., Miskovsky, E., Fields, A.P. and Chapkin, R.S. Quantification of human intestinal gene expression profiles using exfoliated colonocytes: a pilot study. Biomarkers 8: 51-61 (2003).

Koenker, R. Quantile Regression, Econometric Society Monograph Series, Cambridge University Press (2005).

Landis, S. H., Murray, T., Bolden S. and Wingo, P.A. "Cancer Statistics." CA Cancer J Clin 48: 6-29 (1998).

Liu, L-Y.D., Wang, N., Lupton, J.R., Turner, N.D., Chapkin, R.S. and Davidson, L.A. A two-stage normalization method for partially degraded mRNA microarray data. Bioinformatics (Submitted, 2005).

Rodionov, O. N. and Potapov, M.G. Cancer Incidence and Mortality of Pilots and Workers of Avia-Technical Center (An Example of Aeroflot Airline). Medizin und Mobilitat, Fliegerhorst Furstenfeldbruck, Germany (2002).

Williams, J., Zhang, H., Zhou, H., Osman, M., Cha, D., Kave, R., Cuccinotta, F., Dicello, J. and Dillehay, L. "Predicting cancer rates in astronauts from animal carcinogenesis studies and cellular markers." Mutation Res 430: 225-269 (1999).

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Research, 30(4):e15 (2002).

Appendix

Submitted manuscripts

Manuscripts dealing with the isolation and processing of fecal polyA will be submitted once the patent application has been completed and officially filed.

Liu, L.D., N. Wang, J.R. Lupton, N.D. Turner, R.S. Chapkin, and L.A. Davidson. 2005. A two-stage normalization method for partially degraded mRNA microarray data. Bioinformatics (Submitted).

Abstracts

Turner, N.D., L.A. Davidson, L.A. Braby, J.R. Ford, R.J. Carroll, N. Wang, R.S. Chapkin and J.R. Lupton. 2005. Global transcriptional profiling using fecal material as a non-invasive biomarker of colon carcinogenesis. Bioastronautics Investigators' Workshop. January 10-12, 2005, Galveston, TX.

Lupton, J.R., L.M. Sanders, J.C. Mann, N. Popovic, A.A. Glagolenko, L.A. Davidson, L.A. Braby, J.R. Ford, R.J. Carroll, N. Wang, R.S. Chapkin, and N.D. Turner. 2005. A combination of omega-3 fatty acids and a butyrate-producing fiber mitigates colon cancer development. 16th Annual NASA Space Radiation Investigators' Workshop, Port Jefferson, NY, May 15-18, 2005.